

Capacity Achieving Linear Codes with Random Binary Sparse Generating Matrices

A. Makhdoumi Kakhaki, H. Karkeh Abadi, P. Pad, H. Saeedi, F. Marvasti, K. Alishahi

Advanced Communications Research Institute, Sharif University of Technology, Tehran, Iran

Email: {makhdoumi, karkehabadi, pedram_pad}@ee.sharif.edu,

hsaeedi@ieee.org, marvasti@sharif.edu, alishahi@sharif.edu

Abstract

In this paper, we prove the existence of capacity achieving linear codes with random binary sparse generating matrices. The results on the existence of capacity achieving linear codes in the literature are limited to the random binary codes with equal probability generating matrix elements and sparse parity-check matrices. Moreover, the codes with sparse generating matrices reported in the literature are not proved to be capacity achieving.

As opposed to the existing results in the literature, which are based on optimal maximum a posteriori decoders, the proposed approach is based on a different decoder and consequently is suboptimal. We also demonstrate an interesting trade-off between the sparsity of the generating matrix and the error exponent (a constant which determines how exponentially fast the probability of error decays as block length tends to infinity). An interesting observation is that for small block sizes, less sparse generating matrices have better performances while for large block sizes, the performance of the random generating matrices become independent of the sparsity. Moreover, we prove the existence of capacity achieving linear codes with a given (arbitrarily low) density of ones on rows of the generating matrix. In addition to proving the existence of capacity achieving sparse codes, an important conclusion of our paper is that for a sufficiently large code length, no search is necessary in practice to find a deterministic matrix by proving that any arbitrarily selected sequence of sparse generating matrices is capacity achieving with

high probability. The focus in this paper is on the binary symmetric and binary erasure channels.

I. INTRODUCTION

The Shannon coding theorem [1] states that for a variety of channels with a given capacity C , if the information transmission rate R over the channel is below C , there exists a coding scheme for which the information can be transmitted with an arbitrarily low probability of error. For Discrete Memoryless Channels (DMC), it has been shown [2] that the probability of error can be bounded between two exponentially decaying functions of the codeword block length, n . In this theorem, there is no constraint on the codes in terms of linearity. In [3], a simpler proof of the Shannon theorem has been provided. The existence of capacity achieving linear codes over the Binary Symmetric Channel (BSC) was shown by Elias [5] where it was also proved that linear codes have the same error exponent as random codes. A similar result has been obtained in [6]. It was recently shown in [7] that the error exponent of a typical random linear code can, in fact, be larger than a typical random code, implying a faster decaying of error as n increases. Some bounds on the decoding error probability of linear codes have been derived in [8]. The result reported in [5]-[8] are all based on the fact that the elements of generating matrices of the capacity achieving linear codes should be one or zero with equal probability; therefore the generating matrix of such approaches are not sparse.¹ Moreover, most papers on capacity achieving sparse linear codes are concentrated on codes with sparse parity-check matrices. In particular, an important class of codes called Low-Density Parity-Check (LDPC) codes [9], [10] have been of major interest in the past decade. While these codes have sparse parity-check matrices, they do not necessarily exhibit sparse generating matrices which are the focus of this paper. In [11]-[12], some Low-Density Generating-Matrix (LDGM) schemes have been proposed which have performance approaching the capacity.² Some other related literature on

¹A sparse generating matrix is a matrix with a statistically low density of ones, see Section II for the exact definition.

²We distinguish between “capacity approaching” and “capacity achieving” codes. The former term is used when the performance of the code can be shown numerically to approach capacity without any guarantee to achieve it. The latter term is used if the performance can be rigorously proved to achieve the capacity. The subject of this paper is on the latter case.

the codes with sparse generating matrices having performance close to capacity includes [13]-[15]; in [13], a capacity-achieving scheme has been proposed based on serially concatenated codes with an outer LDPC code and an inner LDGM code. However, the generating matrix corresponding to the concatenation is not necessarily sparse. On the other hand, rateless codes have been proposed in [14] and [15] which have sparse generating matrices but are only proved to be capacity achieving over the Binary Erasure Channel (BEC).

In this paper, using a novel approach, we prove the existence of capacity achieving linear codes with *sparse generating matrices* that can provide reliable communications over two important classes of DMC channels; namely, BEC and BSC at rates below the channel capacity. The proof is accomplished by first deriving a lower bound on the probability of correct detection for a given generating matrix and then by taking the expectation of that lower bound over all possible generating matrices with elements 1 and 0 with probability ρ and $1 - \rho$, respectively. By showing that this expectation goes to one as n approaches infinity, we prove the existence of linear capacity achieving codes. To show the sparsity, we extend this result by taking the expectation over a subset of matrices for which the density of ones could be made arbitrarily close to any target ρ . We then prove a stronger result that indicates the existence of capacity achieving linear codes with the same low density of ones in each row of the generating matrix. In addition to proving the existence of capacity achieving sparse codes, we also show that for a sufficiently large code length, no search is necessary in practice to find the desired deterministic matrix. This means that any randomly chosen code can have the desired error correcting property with high probability. This is done by proving that the error probability of a sequence of codes, corresponding to a randomly selected sequence of sparse generating matrices tends to zero as n approaches infinity, in probability. This important result is then extended to generating matrices with low density rows for the case of BSC.

Although in reality the blocklength of codes is finite, in order to prove that a class of codes is capacity achieving, we assume that the blocklength goes to infinity. An interesting question is that for a given error

probability and blocklength, how close the rate of the code can be to the capacity. An upper bound for the channel coding rate achievable at a given blocklength and error probability is derived in [4]. In our paper we use Yuri's upper bound [4] and other well-known results to compare to our numerically derived results.

An interesting trade-off between the sparsity of the generating matrix and the error exponent is demonstrated such that the sparser the matrix, the smaller the error exponent becomes. It is important to note that for the case of BSC, we rigorously prove the existence of capacity achieving linear codes for a constant ρ resulting in a non-vanishing density of ones on the generating matrix as n tends to infinity. However, we have made a conjecture that if we choose $\rho(n) = 1/n^\gamma$; where $0 < \gamma < 1$, the resulting codes can still be capacity achieving, which implies a vanishing density of ones. This signifies that the number of ones in the generating matrix can be as low as $n^{2-\gamma}$. For the case of BEC, we have been able to prove that to have capacity achieving generating matrices, $\rho(n)$ can be of $O(\frac{\log n}{n})$. This implies that the number of ones in the generating matrix is about $n \log n$ which is asymptotically less than $n^{2-\gamma}$, the number of ones in the case of BSC. As opposed to the existing results in the literature, which are based on Maximum A Posteriori (MAP) decoders, the proposed proofs are based on a suboptimal decoder,³ which makes our approach also novel from decoder point of view.

The organization of the paper is as follows: In the next section, some preliminary definitions and notations are presented. In Sections III and IV, we present our theorems for BSC and BEC, respectively, and Section V concludes the paper.

II. PRELIMINARIES

Consider a DMC which is characterized by \mathcal{X} and \mathcal{Y} as its input and output alphabet sets, respectively, and the transition probability function $\mathbb{P}(y|x)$, where $x \in \mathcal{X}$ is the input, and $y \in \mathcal{Y}$ is the output of the channel. In this paper, we consider the binary case where $\mathcal{X} = \{0, 1\}$. A binary code $\mathcal{C}(n, k)$ of rate R is a

³See the details in the next section.

mapping from the set of 2^k k -tuples X_i to n -tuples Z_i , $0 \leq i \leq 2^k - 1$, where $X_i \in \{0, 1\}^k$, $Z_i \in \{0, 1\}^n$, and the code rate R is defined as the ratio of k by n . Since we are only interested in *Linear Codes*, the mapping is fully specified by an $n \times k$ binary matrix $\mathbf{A} = \{A_{ij}\}$ (the generating matrix), and encoding is accomplished by a left multiplication by \mathbf{A} :

$$Z_i = \mathbf{A}X_i,$$

where the calculations are in $\mathbb{GF}(2)$. The vector Z_i is then transmitted through the DMC. Decoding is defined as recovering the vector X_i from the possibly corrupted received version of Z_i .

In this paper the employed decoding scheme relies on the a posteriori probability distribution. Let \mathbf{A} be the generating matrix. For a received vector $Y = y$, the decoder allocates a random vector such as $X = x$ as the original transmitted message with the conditional probability $\mathbb{P}(X = x|Y = y)$. Clearly, the probability of correct detection using \mathbf{A} as the generating matrix is

$$p_c(\mathbf{A}) = \sum_{i,j} \mathbb{P}(x_i) \mathbb{P}(y_j|x_i) \mathbb{P}(x_i|y_j) = \sum_{i,j} \mathbb{P}(x_i, y_j) \mathbb{P}(x_i|y_j) = \mathbb{E}_{X,Y}(\mathbb{P}(X|Y)), \quad (1)$$

where $\mathbb{P}(X, Y)$ depends on \mathbf{A} .

Note that the optimal decoder is a MAP decoder which allocates $\argmax_x \mathbb{P}(X = x|Y = y)$ and that the probability of correct detection using MAP is more than or equal to the probability of correct detection in (1). Throughout the paper, the index i in X_i and Z_i may be dropped for more clarity. For the sake of convenience, the following notations are used for the remainder of the paper.

Definition 1: Let $\mathcal{A}_{n \times k}$ be the set of all binary $n \times k$ matrices. The density of an $\mathbf{A} \in \mathcal{A}_{n \times k}$ is defined as the total number of ones within the matrix divided by the number of its elements (nk). A matrix with a density less than 0.5 is called sparse; the smaller the density, the sparser the matrix becomes.

Definition 2: Let each entry of each element of $\mathcal{A}_{n \times k}$ has a Bernoulli(ρ) distribution, $0 < \rho < 1$.⁴ This scheme induces a *probability distribution* on the set $\mathcal{A}_{n \times k}$, denoted by Bernoulli(n, k, ρ). For the rest of paper, we consider this distribution on the set $\mathcal{A}_{n \times k}$.

⁴A binary random variable has Bernoulli(ρ) distribution if it is equal to 1 with probability of ρ and equal to 0 with probability of $1 - \rho$.

Note that as n approaches infinity, the typical matrices of $\mathcal{A}_{n \times k}$ have a density close to ρ .

III. BINARY SYMMETRIC CHANNEL (BSC)

Consider a BSC with cross-over probability ϵ . The capacity of this channel is given by $C = 1 - h(\epsilon)$, where $h(\epsilon) = -\epsilon \log \epsilon - (1 - \epsilon) \log (1 - \epsilon)$. We suppose that R , the rate of the code, is less than C . In this section, we prove the existence of capacity achieving linear codes with arbitrarily sparse generating matrices over the BSC. We prove the existence by showing that the average error probability over such generating matrices tends to zero as n approaches infinity.

A. Channel Modeling

Assume that we encode a message vector X to generate the codeword $\mathbf{A}X$. Note that X is chosen uniformly from the set $\{0, 1\}^k$. Due to the effect of error in the BSC, each entry of the transmitted codeword $\mathbf{A}X$ can be changed from 0 to 1 and vice versa. These changes can be modeled by adding 1 to erroneous entries of $\mathbf{A}X$ (in $\mathbb{GF}(2)$). Therefore, the error of a BSC with cross-over probability ϵ can be modeled by a binary n -dimensional error vector N with i.i.d. entries with Bernoulli(ϵ) distribution. Thus, if the output of the channel is shown by Y , the following equation models the channel:

$$Y_{n \times 1} = \mathbf{A}_{n \times k} X_{k \times 1} + N_{n \times 1}. \quad (2)$$

Note that X and N are independent.

B. Capacity achieving sparse linear codes for the BSC

In the following theorem, a lower bound for the average probability of correct detection over the set $\mathcal{A}_{n \times k}$, is obtained.

Theorem 1: Consider a BSC with cross-over probability ϵ . A lower bound for the average probability of correct detection over all $n \times k$ generating matrices with Bernoulli(n, k, ρ) distribution is given by

$$\mathbb{E}_{\mathbf{A} \in \mathcal{A}_{n \times k}} (p_c(\mathbf{A})) \geq \sum_{i=0}^n \binom{n}{i} \times \frac{2^n \epsilon^{2i} (1 - \epsilon)^{2(n-i)}}{\sum_{j=0}^k \binom{k}{j} (1 - (1 - 2\epsilon)(1 - 2\rho)^j)^i (1 + (1 - 2\epsilon)(1 - 2\rho)^j)^{n-i}}. \quad (3)$$

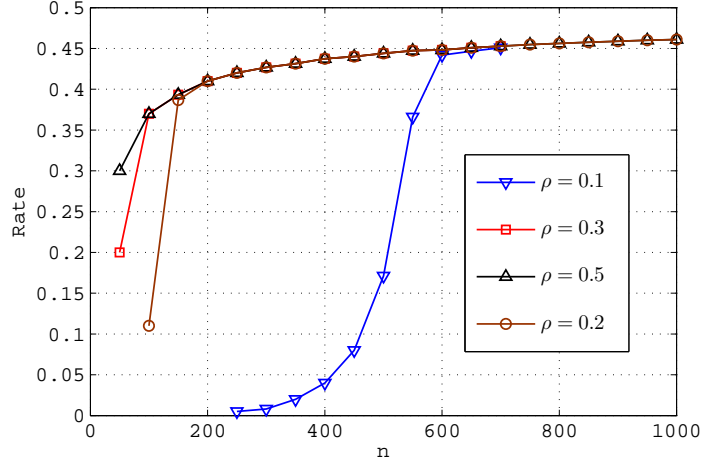


Fig. 1. A comparison of the coding rate versus the blocklength for various values of density(ρ), $\epsilon = 0.11$, Capacity=0.5, error probability= 10^{-1} .

Proof: See Appendix II.

Note 1: An important result of this theorem is that we can fix the error probability and find the maximal achievable rate for a given blocklength. See the following figures.

Figure 1 is a plot of the coding rate versus n for ρ equal to 0.1, 0.3 and 0.5. This plot is numerically evaluated from Theorem 1. An interesting observation of this figure is that when the blocklength n increases, the coding rate becomes independent of the density ρ . This observation can be shown to be true from (22) of Lemma 2, where the parameter ρ disappears on the right hand side. The significance of this observation is that sparse generating matrices can replace non-sparse ones for large block coding sizes, which implies simpler encoder design. This observation is the dual of LDPC codes where large sparse parity check matrices simplifies the decoder design, while the performance remains the same.

Figure 2 is a comparison of our result to that of Gallager result and Yuri upper bound [4]. This figure shows that our results are within the Yuri upper bound and the Gallager result. This figure also shows that for the probability of error equal to 10^{-3} when n becomes greater than 180, the performance of the sparse generating matrices with $\rho = 0.3$ becomes the same as the non-sparse matrices with $\rho = 0.5$.

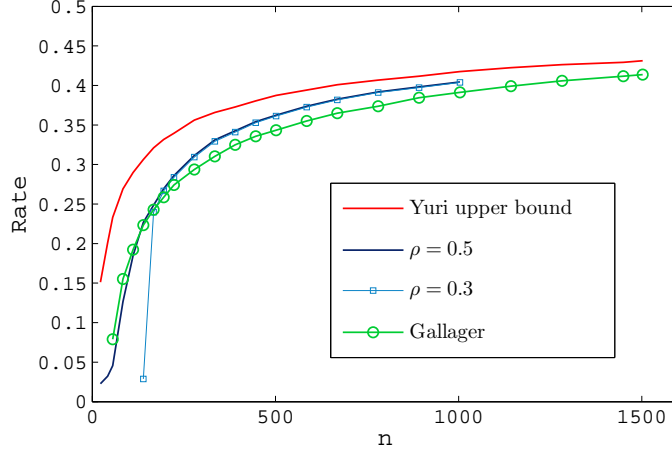


Fig. 2. A comparison of the coding rate versus the blocklength for various methods; the Gallager and the Yuri curves are plotted from [4], $\epsilon = 0.11$, Capacity=0.5, error probability= 10^{-3} .

In the following theorem, we will show that the expected value of the correct detection probability over all generating matrices from $\mathcal{A}_{n \times k}$ approaches 1. This proves the existence of at least one linear capacity achieving code.

Theorem 2: For any $0 < \rho < 1$, for a BSC we have

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbf{A} \in \mathcal{A}_{n \times k}}(p_c(\mathbf{A})) = 1. \quad (4)$$

Proof: See Lemmas 1 and 2 and the proof in Appendix III.

The performance of linear codes is determined by the error exponent which is defined as follows:

Definition 3: The error exponent of a family of codes \mathcal{C} of rate R is defined as

$$E_C(R) = \lim_{n \rightarrow \infty} -\frac{1}{n} \log p_e, \quad (5)$$

where p_e is the average probability of decoding error.

If the limit is greater than zero, the average error probability of the proposed codes decreases exponentially to zero as n increases. The error exponent is an index such that the larger the error exponent,

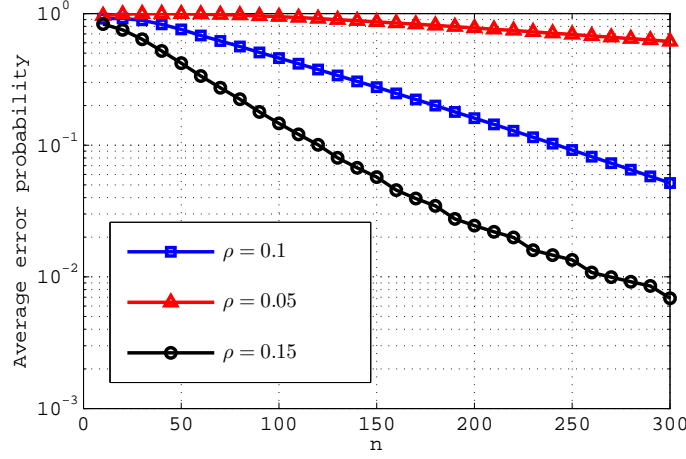


Fig. 3. The average error probability versus n , $\epsilon = 0.05$, $R = 0.8C$

the faster the probability of error decays as n increases. Based on our observation, there is an interesting relation between the error exponent of the codes constructed by generating matrices with Bernoulli(n, k, ρ) distribution and the values of ρ . In Fig. 3, we have plotted the average probability of error versus n for various values of ρ . As it can be seen, the error exponent which is equal to the slope of the curves, increases as ρ increases (the generating matrix become less sparse). In other words, although the probability of error for sparse codes goes to zero exponentially as n increases; this decrease is not as fast as high density codes.

Definition 4: Let $W(A)$ be the number of ones in a given binary matrix A and η be an arbitrary positive constant. $\mathcal{T}_{n \times k}^\eta$ is defined as a subset of $\mathcal{A}_{n \times k}$ for which $|\frac{W(A)}{nk} - \rho| < \eta$, $\eta > 0$. By choosing a sufficiently small η , the set $\mathcal{T}_{n \times k}^\eta$ is in fact a subset of $\mathcal{A}_{n \times k}$ which contains matrices having density of ones arbitrarily close to any given ρ . Note that the probability distribution on $\mathcal{T}_{n \times k}^\eta$ is induced from the probability distribution on $\mathcal{A}_{n \times k}$.

In Theorems 1 and 2, we proved the existence of capacity achieving codes for any value of ρ . We did not explicitly prove the existence of sparse capacity achieving codes. However, using concentration

theory [16], we can see that for a sufficiently large n , a randomly chosen matrix from $\mathcal{A}_{n \times k}$ is in the subset $\mathcal{T}_{n \times k}^\eta$ with high probability. In other words, we can state the following proposition which implies the existence of capacity achieving codes which are sparse.

Proposition 1: Let $\mathcal{T}_{n \times k}^\eta$ be the set of typical matrices defined in Definition (4). We then have

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbf{A} \in \mathcal{T}_{n \times k}^\eta} (p_e) = 0. \quad (6)$$

Definition 5: We define $\mathcal{R}_{n \times k}$ as the set of all binary $n \times k$ matrices with rows that have $k\rho$ ones. We also consider a uniform distribution on the set $\mathcal{R}_{n \times k}$ for the rest of the paper.

In the next theorem, we will prove a stronger result on capacity achieving sparse codes. We show the existence of capacity achieving matrices with rows containing exactly $k\rho$ ones. In other words, the density of ones in each row is exactly equal to ρ . This also implies that the generating matrix has a density of ones exactly equal to ρ . In Theorem 3, we shall derive a lower bound on the average probability of correct detection and in Theorem 4 we will prove that this lower bound tends to one. This shows that the average probability of error over the set $\mathcal{R}_{n \times k}$ approaches zero, implying the existence of capacity achieving codes with generating matrices taken from $\mathcal{R}_{n \times k}$.

Theorem 3: For a binary symmetric channel with cross-over probability ϵ , a lower bound for the expected value of the probability of correct detection over all generating matrices in $\mathcal{R}_{n \times k}$ is given by

$$\mathbb{E}_{\mathbf{A} \in \mathcal{R}_{n \times k}} (p_c(\mathbf{A})) \geq \sum_{i=0}^n \binom{n}{i} \epsilon^i (1 - \epsilon)^{(n-i)} \frac{\epsilon^i (1 - \epsilon)^{(n-i)}}{\sum_{j=0}^k \binom{k}{j} (\epsilon A_j + (1 - \epsilon) B_j)^i ((1 - \epsilon) A_j + \epsilon B_j)^{n-i}}. \quad (7)$$

where

$$A_j = \sum_{q \text{ is odd}} \frac{1}{\binom{k}{k\rho}} \binom{j}{q} \binom{k-j}{k\rho-q}, \quad B_j = \sum_{q \text{ is even}} \frac{1}{\binom{k}{k\rho}} \binom{j}{q} \binom{k-j}{k\rho-q}.$$

Proof: See Appendix IV.

Theorem 4: For each $0 < \rho < 1$, we have

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbf{A} \in \mathcal{R}_{n \times k}} (p_c(\mathbf{A})) = 1. \quad (8)$$

Proof: See Lemma 3 and the proof in Appendix IV.

In Theorems 1 and 2, we proved the existence of capacity achieving linear codes with generating matrices having Bernoulli(n, k, ρ) distribution by showing that the average probability of error over all generating matrices tends to zero as n approaches infinity. This implies that we may have to perform a search over $\mathcal{A}_{n \times k}$ to find such a matrix. Assume that we simply pick matrices randomly for each n from the set $\mathcal{A}_{n \times k}$. This constitutes a sequence of $n \times nR$ matrices. Now consider the resulting sequence of error probabilities corresponding to the sequence of generating matrices. In the following proposition, we shall prove that the limit of this sequence is zero in probability, i.e., a sequence of randomly chosen matrices is capacity achieving with high probability. This suggests that for sufficiently large n , no search is necessary to find a desired deterministic generating matrix.

Proposition 2: Let $\{\mathbf{A}_{n \times nR}\}_{n=0}^{\infty}$ be the sequence of matrices, where $\mathbf{A}_{n \times nR}$ is selected randomly from $\mathcal{A}_{n \times nR}$. If we denote the error probability of the generating matrix $\mathbf{A}_{n \times nR}$ over BSC by $p_e(\mathbf{A}_n)$, then $p_e(\mathbf{A}_n)$ converges in probability to zero as n tends to infinity.

Proof: See Appendix V.

Note 2: If we use the result of Theorem 4, we can extend Proposition 2 to the case where we construct the matrix sequence by choosing the matrices from the set $\mathcal{R}_{n \times k}$. In other words, in order to have capacity achieving sequences of generating matrices for BSC with arbitrarily low density rows, we can simply pick generating matrices randomly from $\mathcal{R}_{n \times k}$.

At this stage, we have been able to rigorously prove the existence of capacity achieving sparse linear codes over the BSC. However for a given ρ , although the density of ones can be made arbitrarily small, it does not go to zero even when n approaches infinity. Let us assume the case where ρ is a decreasing function of n such that $\lim_{n \rightarrow \infty} \rho(n) = 0$, resulting in zero density of ones as n goes to infinity. In the following conjecture, we will propose a result indicating that this assumption can in fact be true. Although, we have not been able to rigorously prove the conjecture, a sketch of the proof has been presented in the

appendix.

Conjecture 1: Let γ be an arbitrary number from interval $(0, 1)$. For $\rho(n) = \frac{1}{n^\gamma}$ by assuming the Bernoulli($n, k, \rho(n)$) distribution on the set $\mathcal{A}_{n \times k}$, we have

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbf{A} \in \mathcal{A}_{n \times k}}(p_c(\mathbf{A})) = 1 \quad (9)$$

See Appendix V for the sketch of the proof.

IV. BINARY ERASURE CHANNEL

A binary erasure channel is identified by erasure probability ϵ and the capacity of this channel is given by $1 - \epsilon$. We use the decoder proposed in Section II. Through the channel, some entries of the coded vector $\mathbf{A}X$, shown by Z , may be erased. According to the position of the erased entries, the error of the channel can be modeled as a subset F of $\mathcal{F} = \{1, \dots, n\}$. Therefore, we employ a decoder which decides about the transmitted vector by observing only the non-erased entries denoted by Z_F . For each $i \in F$, the i^{th} row of \mathbf{A} is removed to derive \mathbf{A}_F . Therefore, the encoding and channel operation can be written as $Z_F = \mathbf{A}_F X$. The decoder chooses \hat{X} , the estimation of X , randomly from the set $\mathcal{X}(Z, F) = \{X | \mathbf{A}_F X = Z_F\}$. In this case, the decoder is equivalent to the MAP decoder. From linear algebra, it can be shown that $|\mathcal{X}(Z, F)| = 2^{k - \text{rank}(\mathbf{A}_F)}$, where rank is the maximum number of independent rows of a matrix calculated in $\mathbb{GF}(2)$. Since \hat{X} is chosen uniformly from $\mathcal{X}(Z, F)$, the probability of the correct detection of X is equal to $2^{-(k - \text{rank}(\mathbf{A}_F))}$. Thus, we have $p_{c|X, F}(\mathbf{A}) = \mathbb{P}(\hat{X} = X | X, F) = 2^{\text{rank}(\mathbf{A}_F) - k}$, where $p_{c|X, F}$ represents the probability of correct detection when X is transmitted and the position of erased entries are given in F .

Theorem 5: Let C be the capacity of a BEC and $\mathbf{A} \in \mathcal{A}_{n \times k}$ is a generating matrix corresponding to a code of rate $R < C$. For any $\rho(n)$ of $O(\frac{\log n}{n})$, the expected value of $p_c(\mathbf{A})$ over all matrices with Bernoulli($n, k, \rho(n)$) distribution tends to 1 as n approaches infinity.

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbf{A} \in \mathcal{A}_{n \times k}}(p_c(\mathbf{A})) = 1. \quad (10)$$

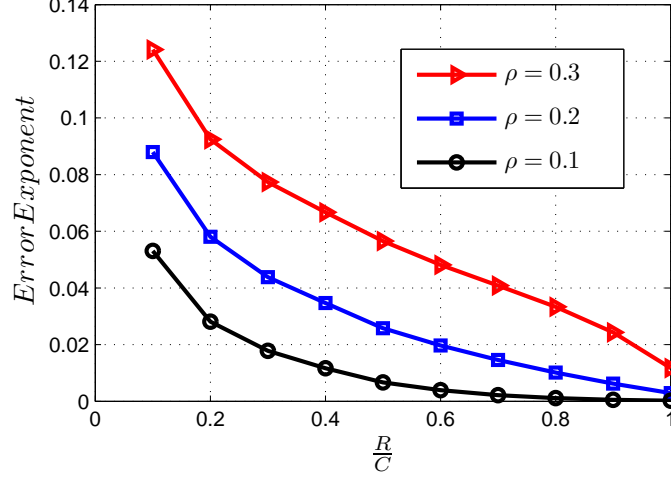


Fig. 4. The error exponent versus $\frac{R}{C}$ for different values of ρ for BEC, $\epsilon = 0.01$

Proof: See Appendix VI.

From the concentration theory [16], similar to the case of the BSC, we can state the following proposition.

Proposition 3: For a BEC with capacity C , codes of rate $R < C$ and generating matrix from $\mathcal{T}_{n \times k}^\eta$, we have:

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbf{A} \in \mathcal{T}_{n \times k}^\eta} (p_e(\mathbf{A})) = 0. \quad (11)$$

In the following proposition we show that similar to Proposition 2 for BSC, a sequence of randomly chosen generating matrices from $\mathcal{A}_{n \times nR}$, results in a capacity achieving coding scheme with high probability. This suggests that for sufficiently large n , no search is necessary to find a desired deterministic generating matrix.

Proposition 4: Let $\{\mathbf{A}_{n \times nR}\}_{n=0}^\infty$ be the sequence of matrices, where $\mathbf{A}_{n \times nR}$ is selected randomly from $\mathcal{A}_{n \times nR}$. If we denote the error probability of the generating matrix $\mathbf{A}_{n \times nR}$ over BEC by $p_e(\mathbf{A}_n)$, then $p_e(\mathbf{A}_n)$ converges in probability to zero as n tends to infinity.

Proof: The proof is similar to the proof of Proposition 2 and thus omitted.

In Fig. 2, we have shown the error exponent as a function of $\frac{R}{C}$ for different values of ρ . As it can be seen, a similar trade-off to BSC exists between sparsity and the error exponent. The smaller ρ results in a smaller error exponent.

The following theorem is similar to Theorem 4.

Theorem 6: For each $0 < \rho < 1$, for a BEC we have

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbf{A} \in \mathcal{R}_{n \times k}}(p_c(\mathbf{A})) = 1. \quad (12)$$

Proof: See Appendix VII.

Note 3: If we use the result of Theorem 6, we can extend Proposition 4 to the case where we construct the matrix sequence by choosing the matrices from the set $\mathcal{R}_{n \times k}$.

V. CONCLUSIONS

In this paper, a novel approach to prove the existence of capacity achieving sparse linear codes over the BSC and BEC was proposed. For the BSC, in Theorem 1, we derived a lower bound on the average probability of correct detection over the set $\mathcal{A}_{n \times k}$. In Theorem 2, we proved that the average probability of error over $\mathcal{A}_{n \times k}$ tends to zero. Then we proved the existence of sparse capacity achieving codes in Proposition 2. In Theorem 3, we derived a lower bound on the average probability of correct detection over the set $\mathcal{R}_{n \times k}$. Using this lower bound in Theorem 4, we proved the existence of capacity achieving codes with generating matrices with the same density (ρ) in each row. In Proposition 2 and its preceding note, we showed that the error probability of codes corresponding to any randomly chosen sequence of generating matrices tends to zero in probability. This implies that for sufficiently large n , a randomly chosen matrix from $\mathcal{A}_{n \times k}$ and $\mathcal{R}_{n \times k}$ will have the average error correcting capability. In addition, we conjectured that Theorem 2 can hold for the case where ρ is a function of n , i.e. $\rho = 1/n^\gamma$. This implies that for a capacity achieving code over a BSC, the density of the generating matrix can approach zero. In Theorem 5 and Proposition 3, we proved the existence of sparse codes for the case of BEC with

generating matrices having Bernoulli distribution with $\rho(n)$ of $O(\frac{\log n}{n})$. A similar result to Proposition 2 and Theorem 4 was shown for BEC in Proposition 4 and Theorem 6, respectively. We demonstrated an interesting trade-off between the sparsity of the generating matrix and the error exponent indicating that a sparser generating matrix results in a smaller error exponent. We also observed that for small block sizes, generating matrices with higher densities have better performances while for large block sizes, the performance of the random generating matrices become independent of the density. In our proofs, we have used a suboptimal decoder while previous works in the literature were based on a MAP decoder. This implies that we can get stronger results if we use the optimal MAP decoder.

For future work, one can try to rigorously prove Conjecture 1 and possibly extend it to the case of matrices in the set $\mathcal{R}_{n \times k}$. The improvement in the bounds using a MAP decoder can be an interesting topic to investigate. The extension of the results to DMC's is another challenging topic to be explored. A very interesting work is to analytically derive the error exponent to prove the trade-off between error exponent and sparsity of the generating matrix.

APPENDIX A

We need the following definitions in order to prove our theorems.

Definition 6: Any two functions $a(n)$ and $b(n)$ are referred to as *proportionally equivalent* and written as $a(n) \approx b(n)$ if $\lim_{n \rightarrow \infty} \frac{a(n)}{b(n)} = 1$.

Definition 7: Any two functions $c(n)$ and $d(n)$ are referred to as *differentially equivalent* and written as $c(n) \doteq d(n)$ if $\lim_{n \rightarrow \infty} c(n) - d(n) = 0$.

APPENDIX B

THE PROOF OF THEOREM 1

Proof of Theorem 1: According to (1), Bayes' rule, and the independency of X and N , we have

$$p_c(\mathbf{A}) = \mathbb{E}_{X,Y} \left(\frac{\mathbb{P}(X)\mathbb{P}(Y|X)}{\mathbb{P}(Y)} \right) = \mathbb{E}_{X,N} \left(\frac{\mathbb{P}(N)\mathbb{P}(X)}{\mathbb{P}(Y)} \right). \quad (13)$$

Taking expectation over all matrices $\mathbf{A} \in \mathcal{A}_{n \times k}$, we get

$$\mathbb{E}_{\mathbf{A}}(p_c(\mathbf{A})) = \mathbb{E}_{\mathbf{A},X,N} \left(\frac{\mathbb{P}(N)\mathbb{P}(X)}{\mathbb{P}(Y)} \right) = \mathbb{E}_{X,N} \left(\mathbb{P}(N)\mathbb{P}(X) \mathbb{E}_{\mathbf{A}} \left(\frac{1}{\mathbb{P}(Y)} \right) \right), \quad (14)$$

where in the last equality, the independency among \mathbf{A} , N and X is used. Using the Jensen's inequality (see [17], Chapter 2, Page 25), we have

$$\begin{aligned} \mathbb{E}_{\mathbf{A}}(p_c(\mathbf{A})) &\geq \mathbb{E}_{X,N} \left(\frac{\mathbb{P}(N)\mathbb{P}(X)}{\mathbb{E}_{\mathbf{A}}(\mathbb{P}(Y))} \right) \\ &= \mathbb{E}_{X,N} \left(\frac{\mathbb{P}(N)\mathbb{P}(X)}{\mathbb{E}_{\mathbf{A}}(\mathbb{P}_{X',N'}(\mathbf{A}X + N = \mathbf{A}X' + N'))} \right) \\ &= \mathbb{E}_{X,N} \left(\frac{\mathbb{P}(N)\mathbb{P}(X)}{\mathbb{E}_{\mathbf{A}}(\mathbb{E}_{X',N'}(1_{[\mathbf{A}(X-X')+(N-N')=0]}))} \right) \\ &= \mathbb{E}_{X,N} \left(\frac{\mathbb{P}(N)\mathbb{P}(X)}{\mathbb{E}_{X'}(\mathbb{P}_{\mathbf{A},N'}(\mathbf{A}(X-X') + (N-N') = 0))} \right), \end{aligned} \quad (15)$$

where X' and N' have the same distributions as the input and error vectors, respectively. In the above equation, the expected value over X' is a function of binary subtraction $X - X'$ and as a result does

not depend on X . Thus we can assume any binary vector X such as the all zero vector, X_0 ; from the independency of the rows of \mathbf{A} in (15) and the uniformity of the vectors X , we have

$$\begin{aligned}\mathbb{E}_{\mathbf{A}}(p_c(\mathbf{A})) &\geq \frac{1}{2^k} \mathbb{E}_{N, X=X_0} \left(\frac{\mathbb{P}(N)}{\mathbb{E}_{X'} \left(\prod_{l=1}^n \mathbb{P}_{A_l, N'_l} (A_l(X_0 - X') + (N_l - N'_l) = 0) \right)} \right) \\ &= \frac{1}{2^k} \mathbb{E}_N \left(\frac{\mathbb{P}(N)}{\mathbb{E}_{X'} \left(\prod_{l=1}^n \mathbb{P}_{A_l, N'_l} (A_l(X') + (N_l - N'_l) = 0) \right)} \right),\end{aligned}\quad (16)$$

where N_l and N'_l are the l^{th} entry of N and N' , respectively, and A_l is the l^{th} row of \mathbf{A} . Note that here all the operations are performed in $\mathbb{GF}(2)$. In order to evaluate the right side of the above inequality, assume that vector N has i ones. Without loss of generality and for convenience, we assume that the first i elements of N are 1. Thus, the argument of the expected value in (16) is equal to

$$\frac{\epsilon^i (1 - \epsilon)^{n-i}}{\mathbb{E}_{X'} \left(\prod_{l=1}^i (\mathbb{P}_{A_l, N'_l} (A_l X' + N'_l = 1)) \prod_{l=i+1}^n (\mathbb{P}_{A_l, N'_l} (A_l X' + N'_l = 0)) \right)}.\quad (17)$$

To evaluate the expected value in the above expression, note that since $N'_l = 0$ with probability $1 - \epsilon$ and $N'_l = 1$ with probability ϵ , we have

$$\prod_{l=1}^i (\mathbb{P}_{A_l, N'_l} (A_l X' + N'_l = 1)) = (\epsilon \times \mathbb{P}(A_l X' = 0) + (1 - \epsilon) \times \mathbb{P}(A_l X' = 1))^i.\quad (18)$$

Now assume j elements of X' are equal to 1. Also consider the entries of A_l with the same indices as the entries of X' that are equal to one. It is easy to see that in the above equation, $\mathbb{P}(A_l X' = 1)$ is equal to the probability of having an odd number of ones in the considered indices of A_l . Thus, we have

$$\mathbb{P}(A_l X' = 1) = \sum_{q \text{ odd}} \binom{j}{q} \rho^q (1 - \rho)^{j-q} = \frac{((1 - \rho) + \rho)^j - ((1 - \rho) - \rho)^j}{2} = \frac{1 - (1 - 2\rho)^j}{2}.\quad (19)$$

The same argument results in $\mathbb{P}(A_l X' = 0) = \frac{1 + (1 - 2\rho)^j}{2}$ and therefore we have

$$\prod_{l=1}^n \mathbb{P}_{A_l, N'_l} (A_l(X_0 - X') + (N_l - N'_l) = 0) = \left(\frac{1 - (1 - 2\epsilon)(1 - 2\rho)^j}{2} \right)^i \left(\frac{1 + (1 - 2\epsilon)(1 - 2\rho)^j}{2} \right)^{n-i}.\quad (20)$$

The expectation of the above expression over X' results in

$$\mathbb{E}_{X'} \left(\prod_{l=1}^n \mathbb{P}_{A_l, N'_l} (A_l(X_0 - X') + (N_l - N'_l) = 0) \right) = \sum_{j=0}^k \frac{1}{2^k} \binom{k}{j} \left(\frac{1 - (1 - 2\epsilon)(1 - 2\rho)^j}{2} \right)^i \left(\frac{1 + (1 - 2\epsilon)(1 - 2\rho)^j}{2} \right)^{n-i}. \quad (21)$$

Substituting (21) in (17) and taking expected value with respect to N , we obtain the following lower bound for $\mathbb{E}_{\mathbf{A}}(p_c)$:

$$\mathbb{E}_{\mathbf{A} \in \mathcal{A}_{n \times k}}(p_c(\mathbf{A})) \geq \sum_{i=0}^n \binom{n}{i} \epsilon^i (1 - \epsilon)^{n-i} \frac{2^n \epsilon^i (1 - \epsilon)^{n-i}}{\sum_{j=0}^k (1 - (1 - 2\epsilon)(1 - 2\rho)^j)^i (1 + (1 - 2\epsilon)(1 - 2\rho)^j)^{n-i}}.$$

This completes the proof. ■

APPENDIX C

LEMMA 1, 2 AND THE PROOF OF THEOREM 2

Lemma 1: Let $\{a_i\}_{i=0}^{\infty}$ be a bounded sequence. For any $\delta > 0$ and $0 \leq p \leq 1$ the summation $\sum_{i=0}^n \binom{n}{i} p^i (1 - p)^{n-i} a_i$ is differentially equivalent to $\sum_{i=n(p-\delta)}^{n(p+\delta)} \binom{n}{i} p^i (1 - p)^{n-i} a_i$.

Proof: According to the Chernoff-Hoeffding Theorem [19] the proof is straightforward. ■

Lemma 2: Consider a code with rate R over a BSC with cross-over probability of ϵ where $R = k/n < C = 1 - h(\epsilon)$. There exists a $\delta > 0$ for which for any $i \in (n(\epsilon - \delta), n(\epsilon + \delta))$, we have

$$\sum_{j=0}^k \binom{k}{j} (1 - (1 - 2\epsilon)(1 - 2\rho)^j)^i (1 + (1 - 2\epsilon)(1 - 2\rho)^j)^{n-i} \approx 2^n \epsilon^i (1 - \epsilon)^{n-i}. \quad (22)$$

Proof: To prove the lemma, note that the first term of the summation ($j = 0$) in the left hand side of (22) is equal to the right hand side. Therefore, to prove (22), it is sufficient to show that

$$\sum_{j=1}^k \binom{k}{j} \left(\frac{1 - (1 - 2\epsilon)(1 - 2\rho)^j}{2\epsilon} \right)^i \left(\frac{1 + (1 - 2\epsilon)(1 - 2\rho)^j}{2(1 - \epsilon)} \right)^{n-i} \doteq 0. \quad (23)$$

Let $b(j) = \frac{1-(1-2\epsilon)(1-2\rho)^j}{2}$ and $a(j, i) = \left(\left(\frac{1-(1-2\epsilon)(1-2\rho)^j}{2\epsilon} \right)^{\frac{i}{n}} \left(\frac{1+(1-2\epsilon)(1-2\rho)^j}{2(1-\epsilon)} \right)^{\frac{n-i}{n}} \right)^{\frac{1}{R}}$. Thus, we have

$$a(j, i) = \left(\left(\frac{b(j)}{\epsilon} \right)^{\frac{i}{n}} \left(\frac{1-b(j)}{1-\epsilon} \right)^{\frac{n-i}{n}} \right)^{\frac{1}{R}}. \quad (24)$$

By using a straightforward calculation, it can be shown that for $i = n\epsilon$, the maximum of $a(j, i)$ is equal to 1. The maximum of $a(j, i)$ occurs for $b(j) = \epsilon$ or equivalently $j = 0$. Thus, for $j \geq 1$ we have

$$a(j, n\epsilon) < 1. \quad (25)$$

Also, since $\lim_{j \rightarrow \infty} (b(j)) = \frac{1}{2}$, we have

$$\lim_{j \rightarrow \infty} a(j, n\epsilon) = \left(\left(\frac{1}{2\epsilon} \right)^{\epsilon} \left(\frac{1}{2(1-\epsilon)} \right)^{1-\epsilon} \right)^{\frac{1}{R}} = 2^{-\frac{C}{R}} < \frac{1}{2}. \quad (26)$$

It is easy to see that $a(j, i)$ is a uniformly continuous function of i and j . Thus from (25), we conclude that there is a $\delta_1 > 0$ for which for any $i \in (n(\epsilon - \delta_1), n(\epsilon + \delta_1))$ and $j \geq 1$, we have $a(j, i) < 1$. And also from (26), we conclude that there is a $\delta_2 > 0$ for which for any $i \in (n(\epsilon - \delta_2), n(\epsilon + \delta_2))$, we have $\lim_{j \rightarrow \infty} a(j, i) < \frac{1}{2}$. Let $\delta = \min(\delta_1, \delta_2)$ and fix $i \in (n(\epsilon - \delta), n(\epsilon + \delta))$; there exist an integer M and a real number $\mu > 0$, for which we have $a(j, i) < \frac{1}{2} - \mu$ for all $j > M$. By using this M , the left hand side of (23) can be written as

$$\sum_{j=1}^k \binom{k}{j} a(j, i)^k = \sum_{j=1}^M \binom{k}{j} a(j, i)^k + \sum_{j=M+1}^k \binom{k}{j} a(j, i)^k. \quad (27)$$

Since $a(j, i) < \frac{1}{2} - \mu$ for $j > M$, we have

$$\lim_{k \rightarrow \infty} \sum_{j=M+1}^k \binom{k}{j} a(j, i)^k \leq \lim_{k \rightarrow \infty} \left(\sum_{j=M+1}^k \binom{k}{j} \right) \left(\frac{1}{2} - \mu \right)^k \leq \lim_{k \rightarrow \infty} 2^k \left(\frac{1}{2} - \mu \right)^k = 0. \quad (28)$$

Therefore, $\lim_{k \rightarrow \infty} \sum_{j=M+1}^k \binom{k}{j} a(j, i)^k = 0$.

To see that the first term at the right hand side of (27) also tends to zero, let $w = \max_{1 \leq j \leq M} a(j, i) < 1$.

Therefore, we can write

$$\sum_{j=1}^M \binom{k}{j} a(j, i)^k < \left(\sum_{j=1}^M \binom{k}{j} \right) w^k \leq (Mk^M) w^k = Me^{-(vk - M \ln(k))},$$

where $v = -\ln(w) > 0$. Now the right hand side of the above inequality tends to zero because $vk - M \ln(k)$ tends to infinity as k approaches infinity. This proves that the left hand side should also tend to zero. Therefore, both summations at the right hand side of (27) tend to zero. This proves (23) and consequently (22). ■

Proof of Theorem 2:

Let $a_i = \frac{2^n \epsilon^i (1-\epsilon)^{n-i}}{\sum_{j=0}^k \binom{k}{j} (1-(1-2\epsilon)(1-2\rho)^j)^i (1+(1-2\epsilon)(1-2\rho)^j)^{n-i}}$. The first term of the summation of the denominator is equal to the numerator, and the other terms in the summation are positive. Thus, the elements of the sequence $\{a_i\}_{i=0}^n$ are less than 1 and subsequently bounded. Therefore, we can apply Lemma 1. Now note that based on Theorem 1, we have

$$\mathbb{E}_{\mathbf{A} \in \mathcal{A}_{n \times k}}(p_c(\mathbf{A})) \geq \sum_{i=0}^n \binom{n}{i} \epsilon^i (1-\epsilon)^{n-i} \frac{2^n \epsilon^i (1-\epsilon)^{n-i}}{\sum_{j=0}^k (1-(1-2\epsilon)(1-2\rho)^j)^i (1+(1-2\epsilon)(1-2\rho)^j)^{n-i}}.$$

Let δ be as in Lemma 1. Since $\mathbb{E}_{\mathbf{A} \in \mathcal{A}_{n \times k}}(p_c(\mathbf{A})) \leq 1$, to prove the theorem, it is enough to show that the right hand side of the above inequality is differentially equivalent to 1. To see this, we write

$$\begin{aligned} & \sum_{i=0}^n \binom{n}{i} \epsilon^i (1-\epsilon)^{n-i} \frac{2^n \epsilon^i (1-\epsilon)^{n-i}}{\sum_{j=0}^k (1-(1-2\epsilon)(1-2\rho)^j)^i (1+(1-2\epsilon)(1-2\rho)^j)^{n-i}} \doteq \\ & \sum_{i=n(\epsilon-\delta)}^{n(\epsilon+\delta)} \binom{n}{i} \epsilon^i (1-\epsilon)^{n-i} \frac{2^n \epsilon^i (1-\epsilon)^{n-i}}{\sum_{j=0}^k (1-(1-2\epsilon)(1-2\rho)^j)^i (1+(1-2\epsilon)(1-2\rho)^j)^{n-i}} \doteq \quad (29) \\ & \sum_{i=n(\epsilon-\delta)}^{n(\epsilon+\delta)} \binom{n}{i} \epsilon^i (1-\epsilon)^{n-i} \frac{2^n \epsilon^i (1-\epsilon)^{n-i}}{2^n \epsilon^i (1-\epsilon)^{n-i}} \doteq \sum_{i=0}^n \binom{n}{i} \epsilon^i (1-\epsilon)^{n-i} = 1, \end{aligned}$$

where we used Lemma 1 in the first and third equality⁵ and we replaced the summation in the denominator based on Lemma 2. This proves the theorem. ■

APPENDIX D

THE PROOF OF THEOREMS 3 AND 4

Proof of Theorem 3:

We follow steps similar to that of Theorem 1. Equations (13) to (17) in Theorem 1 still hold here. It can

⁵Note that by equality we mean \doteq which is not mathematically precise but we use it throughout the paper for the ease of explanation.

be easily seen that $\mathbb{P}(A_l X' = 1) = A_j$ and $\mathbb{P}(A_l X' = 0) = B_j$. Thus equation (18) is modified as

$$\prod_{l=1}^n \mathbb{P}_{A_l, N'_l} (A_l(X') + (N_l - N'_l) = 0) = (\epsilon A_j + (1 - \epsilon) B_j)^i (\epsilon B_j + (1 - \epsilon) A_j)^{n-i}. \quad (30)$$

The expectation of the above expression over X' results in

$$\begin{aligned} \mathbb{E}_{X'} \left(\prod_{l=1}^n \mathbb{P}_{A_l, N'_l} (A_l(X_0 - X') + (N_l - N'_l) = 0) \right) = \\ \sum_{j=0}^k \frac{1}{2^k} \binom{k}{j} (\epsilon A_j + (1 - \epsilon) B_j)^i (\epsilon B_j + (1 - \epsilon) A_j)^{n-i}. \end{aligned} \quad (31)$$

Substituting (31) into (17) and taking the expectation with respect to N , we obtain

$$\mathbb{E}_{\mathbf{A} \in \mathcal{R}_{n \times k}} (p_c(\mathbf{A})) \geq \sum_{i=0}^n \binom{n}{i} \epsilon^i (1 - \epsilon)^{n-i} \frac{\epsilon^i (1 - \epsilon)^{n-i}}{\sum_{j=0}^k \binom{k}{j} (\epsilon A_j + (1 - \epsilon) B_j)^i (\epsilon B_j + (1 - \epsilon) A_j)^{n-i}}.$$

This completes the proof. ■

Proof of Theorem 4:

First we prove a lemma similar to Lemma 2.

Lemma 3: Suppose that $R = k/n < 1 - h(\epsilon)$. There exists a $\delta > 0$ for which for any $i \in (n(\epsilon - \delta), n(\epsilon + \delta))$, we have

$$\sum_{j=0}^k \binom{k}{j} (\epsilon A_j + (1 - \epsilon) B_j)^i (\epsilon B_j + (1 - \epsilon) A_j)^{n-i} \approx \epsilon^i (1 - \epsilon)^{n-i}. \quad (32)$$

Proof: Let $b(j) = \epsilon A_j + (1 - \epsilon) B_j$ and $a(j, i) = \left(\left(\frac{\epsilon A_j + (1 - \epsilon) B_j}{\epsilon} \right)^{\frac{i}{n}} \left(\frac{\epsilon B_j + (1 - \epsilon) A_j}{1 - \epsilon} \right)^{\frac{n-i}{n}} \right)^{\frac{1}{R}}$. Since $A_j + B_j = 1$, we have

$$a(j, i) = \left(\left(\frac{b(j)}{\epsilon} \right)^{\frac{i}{n}} \left(\frac{1 - b(j)}{1 - \epsilon} \right)^{\frac{n-i}{n}} \right)^{\frac{1}{R}}. \quad (33)$$

By employing the same approach as the proof of Lemma 2, it is sufficient to show that

$$\lim_{j \rightarrow \infty} b(j) = \frac{1}{2}.$$

It is easy to see that $\lim_{j \rightarrow \infty} A_j = \lim_{j \rightarrow \infty} B_j = \frac{1}{2}$. As a result, we have

$$\lim_{j \rightarrow \infty} b(j) = \lim_{j \rightarrow \infty} (\epsilon A_j + (1 - \epsilon) B_j) = \frac{\epsilon}{2} + \frac{1 - \epsilon}{2} = \frac{1}{2}. \quad (34)$$

This completes the proof of lemma. ■

Now to prove this theorem, it is enough to replace the denominator in summation of (7) with the right hand side of (32) according to Lemma 3. ■

APPENDIX E

PROOF OF PROPOSITION 2 AND PROOF SKETCH OF CONJECTURE 1

Proof of Proposition 2:

In order to show that $\lim_{n \rightarrow \infty} p_e(\mathbf{A}_{n \times nR}) = 0$ in probability, we have to show that for any given $\delta > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(p_e(\mathbf{A}_{n \times nR}) > \delta) = 0.$$

For a given $\xi > 0$, define $\beta = \min\{\delta, \xi\}$. According to Theorem 2, we have $\lim_{n \rightarrow \infty} \mathbb{E}(p_e(\mathbf{A}_{n \times nR})) = 0$.

Thus, there exists an N_β for which for any $n > N_\beta$, $\mathbb{E}(p_e(\mathbf{A}_{n \times nR})) < \beta^2$. Therefore, due to the fact that $p_e(\mathbf{A}_{n \times nR}) \geq 0$, for $n > N_\beta$, we obtain $\mathbb{P}(p_e(\mathbf{A}_{n \times nR}) > \beta) < \beta$. Hence, for $n > N_\beta$, since $\beta \leq \delta$, we have

$$\mathbb{P}(p_e(\mathbf{A}_{n \times nR}) > \delta) \leq \mathbb{P}(p_e(\mathbf{A}_{n \times nR}) > \beta) < \beta < \xi.$$

Thus, for $n > N_\beta$, we have

$$\mathbb{P}(p_e(\mathbf{A}_{n \times nR}) > \delta) < \xi,$$

and the proof is complete. ■

Sketch of proof of Conjecture 1:

The lower bound of Theorem 1 still holds for the case where ρ is a function of n where $0 < \rho(n) < 1$. If we can show that for $R = k/n < 1 - h(\epsilon)$, there exists a $\delta > 0$ for which for any $i \in (n(\epsilon - \delta), n(\epsilon + \delta))$, we have

$$\sum_{j=0}^k \binom{k}{j} (1 - (1 - 2\epsilon)(1 - 2\rho(n))^j)^i (1 + (1 - 2\epsilon)(1 - 2\rho(n))^j)^{n-i} \approx 2^n \epsilon^i (1 - \epsilon)^{n-i}. \quad (35)$$

From the approach similar to that of the proof of Theorem 2, the proof will be straightforward. Although, we have numerical evidence suggesting that the above equality holds, we have not been able to prove it rigorously. The rest of the proof is as follows. Let

$$a_i = \frac{2^n \epsilon^i (1 - \epsilon)^{n-i}}{\sum_{j=0}^k \binom{k}{j} (1 - (1 - 2\epsilon)(1 - 2\rho(n))^j)^i (1 + (1 - 2\epsilon)(1 - 2\rho(n))^j)^{n-i}}. \quad (36)$$

Since, the first term of the summation in the denominator is equal to the numerator, the sequence $\{a_i\}_{i=0}^n$ are less than 1 and subsequently bounded. From Lemma 1 we get

$$\sum_{i=0}^n \binom{n}{i} \epsilon^i (1 - \epsilon)^{n-i} a_i \doteq \sum_{i=n(\epsilon-\delta)}^{n(\epsilon+\delta)} \binom{n}{i} \epsilon^i (1 - \epsilon)^{n-i} a_i.$$

Now we have

$$\begin{aligned} \mathbb{E}_{\mathbf{A}}(p_c(\mathbf{A})) &\geq \sum_{i=0}^n \binom{n}{i} \epsilon^i (1 - \epsilon)^{n-i} \frac{2^n \epsilon^i (1 - \epsilon)^{n-i}}{\sum_{j=0}^k (1 - (1 - 2\epsilon)(1 - 2\rho(n))^j)^i (1 + (1 - 2\epsilon)(1 - 2\rho(n))^j)^{n-i}} \\ &\doteq \sum_{i=n(\epsilon-\delta)}^{n(\epsilon+\delta)} \binom{n}{i} \epsilon^i (1 - \epsilon)^{n-i} \frac{2^n \epsilon^i (1 - \epsilon)^{n-i}}{2^n \epsilon^i (1 - \epsilon)^{n-i}} \\ &\doteq \sum_{i=0}^n \binom{n}{i} \epsilon^i (1 - \epsilon)^{n-i} = 1. \end{aligned}$$

In other words, $\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbf{A} \in \mathcal{A}_{n \times k}}(p_c(\mathbf{A})) = 1$ which is the desired result. ■

APPENDIX F

THE PROOF OF THEOREM 5

Proof of Theorem 5:

We first present a lemma from [18].

Lemma 4: Suppose $\delta \geq 0$ and let $\text{Bernoulli}(n(1 + \delta), n, \rho(n))$ be the probability distribution on the $n(1 + \delta) \times n$ matrices where $\rho(n)$ is of $O(\frac{\log n}{n})$. Then $\mathbb{E}(\text{rank}(\mathbf{A}_{n(1+\delta) \times n})) \approx n$.

Since $k/n < 1 - \epsilon$, it can be concluded that there exists a $\delta > 0$ for which $k = n(1 - \epsilon - \delta)$. By using the proposed decoding scheme and by decomposing $p_{c|X}(\mathbf{A})$ according to the position of the erased

entries F , we get

$$p_{c|X}(\mathbf{A}) = \mathbb{P}(\hat{X} = X|X) = \sum_{F \subseteq \mathcal{F}} \mathbb{P}(\hat{X} = X|X, F) \mathbb{P}(F) = \sum_{F \subseteq \mathcal{F}} \epsilon^{|F|} (1 - \epsilon)^{n-|F|} 2^{\text{rank}(\mathbf{A}_F) - k}.$$

Therefore, $p_{c|X}(\mathbf{A})$ is the same for all X 's. Thus $p_c(\mathbf{A}) = p_{c|X}(\mathbf{A})$. By evaluating the expected value of $p_c(\mathbf{A})$ over all matrices and using Jensen inequality, we have

$$\mathbb{E}_{\mathbf{A}}(p_c(\mathbf{A})) = \sum_{i=0}^n \binom{n}{i} \epsilon^i (1 - \epsilon)^{n-i} \mathbb{E}_{\mathbf{A}}(2^{\text{rank}(\mathbf{A}_{(n-i) \times k}) - k}) \geq \sum_{i=0}^n \binom{n}{i} \epsilon^i (1 - \epsilon)^{n-i} 2^{\mathbb{E}_{\mathbf{A}}(\text{rank}(\mathbf{A}_{(n-i) \times k})) - k}.$$

Applying Lemma 1, we obtain

$$\sum_{i=0}^n \binom{n}{i} \epsilon^i (1 - \epsilon)^{n-i} 2^{\mathbb{E}_{\mathbf{A}}(\text{rank}(\mathbf{A}_{(n-i) \times k})) - k} \doteq \sum_{i=n(\epsilon-\theta)}^{n(\epsilon+\theta)} \binom{n}{i} \epsilon^i (1 - \epsilon)^{n-i} 2^{\mathbb{E}_{\mathbf{A}}(\text{rank}(\mathbf{A}_{(n-i) \times k})) - k},$$

where θ is chosen such that $\theta < \delta$. For each $i \in (n(\epsilon - \theta), n(\epsilon + \theta))$, there is an $\alpha \in (-\theta, \theta)$ for which $i = n(\epsilon - \alpha)$. Therefore, $n - i = n(1 - \epsilon + \alpha) > n(1 - \epsilon - \delta) = k$. Now, according to Lemma 4, if we substitute k for $\mathbb{E}(\text{rank}(\mathbf{A}_{(n-i) \times k}))$, as $n \rightarrow \infty$, we can write

$$\begin{aligned} \mathbb{E}_{\mathbf{A}}(p_c(\mathbf{A})) &\geq \sum_{i=n(\epsilon-\theta)}^{n(\epsilon+\theta)} \binom{n}{i} \epsilon^i (1 - \epsilon)^{n-i} 2^{\mathbb{E}(\text{rank}(\mathbf{A}_{(n-i) \times k})) - k} \\ &= \sum_{i=n(\epsilon-\theta)}^{n(\epsilon+\theta)} \binom{n}{i} \epsilon^i (1 - \epsilon)^{n-i} 2^{k \left(\frac{\mathbb{E}(\text{rank}(\mathbf{A}_{(n-i) \times k}))}{k} - 1 \right)} \\ &\doteq \sum_{i=n(\epsilon-\theta)}^{n(\epsilon+\theta)} \binom{n}{i} \epsilon^i (1 - \epsilon)^{n-i} \doteq \sum_{i=0}^n \binom{n}{i} \epsilon^i (1 - \epsilon)^{n-i} = 1. \end{aligned}$$

Therefore, $\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbf{A}}(p_c(\mathbf{A})) = 1$. ■

APPENDIX G

THE PROOF OF THEOREM 6

Proof of Theorem 6:

According to the proof of Theorem 5 it is sufficient to show the following lemma.

Lemma 5: Suppose $\delta \geq 0$ and consider $\mathcal{R}_{n(1-\delta) \times n}$ with its previously defined distribution. Then for $\mathbf{A}_{n(1-\delta) \times n} \in \mathcal{R}_{n(1-\delta) \times n}$ we have $\mathbb{E}(\text{rank}(\mathbf{A}_{n(1-\delta) \times n})) \approx n(1 - \delta)$. Note that here rank is calculated in $\mathbb{GF}(2)$.

Proof: In order to prove the lemma it is sufficient to show that

$$\lim_{n \rightarrow \infty} \mathbb{P}(\text{rank}(\mathbf{A}_{n(1-\delta) \times n}) = n(1-\delta)) = 1,$$

which is equivalent to show that the probability of having a matrix $\mathbf{A}_{n(1-\delta) \times n}$ with linear dependent rows goes to zero as n approaches infinity, i.e.,

$$\lim_{n \rightarrow \infty} \left(\sum_{k=1}^{n(1-\delta)} \binom{n(1-\delta)}{k} \mathbb{P}(A_1 + A_2 + \dots + A_k = 0) \right) = 0, \quad (37)$$

where A_i represents the i^{th} row of the matrix. Suppose ζ be a positive number such that $\zeta < \rho$. The summation of equation (37) can be written as

$$\sum_{k=1}^{n\zeta} \binom{n(1-\delta)}{k} \mathbb{P}(A_1 + A_2 + \dots + A_k = 0) + \sum_{k=n\zeta}^{n(1-\delta)} \binom{n(1-\delta)}{k} \mathbb{P}(A_1 + A_2 + \dots + A_k = 0). \quad (38)$$

We first prove that the first term tends to zero. In order to have $A_1 + A_2 + \dots + A_k = 0$, A_k should be equal to the sum of A_1 to A_{k-1} . Thus, conditioning on A_1 to A_{k-1} , it is easy to see that $\mathbb{P}(A_1 + A_2 + \dots + A_k = 0) \leq \frac{1}{\binom{n}{\rho n}}$. Thus,

$$\sum_{k=1}^{n\zeta} \binom{n(1-\delta)}{k} \mathbb{P}(A_1 + A_2 + \dots + A_k = 0) \leq \left(\sum_{k=1}^{n\zeta} \binom{n}{k} \right) \frac{1}{\binom{n}{\rho n}}.$$

Since $\binom{n}{n\rho} \approx \frac{2^{h(\rho)n}}{(\rho(1-\rho)n2\pi)^{1/2}}$, we have

$$\lim_{n \rightarrow \infty} \left(\sum_{k=1}^{n\zeta} \binom{n}{k} \right) \frac{1}{\binom{n}{\rho n}} \leq \lim_{n \rightarrow \infty} \frac{n\zeta 2^{nh(\zeta)} 2^{-nh(\rho)}}{(\zeta(1-\zeta))^{1/2} (\rho(1-\rho))^{-1/2}} = 0,$$

where in the last equality we used the fact that $h(\zeta) < h(\rho)$.

in order to complete the proof it is sufficient to show that the second term of equation (38) also goes to zero. In this regard we show that for sufficiently large n and any $k \geq \zeta n$ we have

$$\mathbb{P}(A_1 + A_2 + \dots + A_k = 0) \leq \left(\frac{1}{2^{1-\frac{\delta}{2}}} \right)^n. \quad (39)$$

This will prove the lemma because we would have

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{k=n\zeta}^{n(1-\delta)} \binom{n(1-\delta)}{k} \mathbb{P}(A_1 + A_2 + \dots + A_k = 0) &\leq \\ \lim_{n \rightarrow \infty} \left(\sum_{k=n\zeta}^{n(1-\delta)} \binom{n(1-\delta)}{k} \right) \left(\frac{1}{2^{1-\frac{\delta}{2}}} \right)^n &\leq \lim_{n \rightarrow \infty} 2^{n(1-\delta)} \left(\frac{1}{2^{1-\frac{\delta}{2}}} \right)^n = 0. \end{aligned}$$

In order to prove equation (39) we employ coupling method from random walk theory. Consider a random walk on the n -dimensional cubic in $\mathbb{GF}(2)$ with the set of directions S , consists of all n -dimensional vectors with ρn ones. Suppose this random walk starts from the origin, and each time selects its next direction randomly from S with uniform distribution. Therefore, $\mathbb{P}(A_1 + A_2 + \dots + A_k = 0)$ represents the probability of returning back to the origin after k steps. Denote this random walk by the sequence $\{X_t\}$ of n -dimensional vectors where X_t represents the position of the random walk after t steps. Note that the stationary distribution of this random walk is uniform distribution, which means as t tends to infinity the probability of being at any points of the cubic is almost $(\frac{1}{2})^n$. Thus, for large values of k , $\mathbb{P}(A_1 + A_2 + \dots + A_k = 0)$ is almost $(\frac{1}{2})^n$. Now Consider another random walk denoted by $\{Y_t\}$, which its starting point is selected randomly with the uniform distribution. The idea of coupling is to couple two random walks $\{X_t\}$ and $\{Y_t\}$ with the dependency between the directions selected by them such that both of them remain random walks that select their directions in each step uniformly from S . Suppose X_i and Y_i are the positions of the two random walks after i steps and s_{i+1}^x be the $(i+1)^{th}$ direction which is selected uniformly from S by the random walk $\{X_t\}$. Suppose r_i entries of the vectors X_i and Y_i are the same and denote the positions of these entries by the set $U = \{u_1, u_2, \dots, u_{r_i}\}$. Let S_U be the subset of S consists of vectors that their r_i entries with positions from U are same as s_{i+1}^x . The random walk $\{Y_t\}$ select the direction s_{i+1}^y uniformly from the set S_U . Note that due to the fact that $\{Y_t\}$ starts with its stationary distribution the probability of being at any point remains uniform for all t for this random walk. Also note that according to the dependency between s_{i+1}^x and s_{i+1}^y , $\{r_i\}$ is a non-decreasing sequence. Thus, we expect that the two random walk meet each other at a point. Let τ be the first time that $\{X_t\}$ and $\{Y_t\}$ meet. Note that after τ the rest of the two random walks would be the same. Conditioning on the τ , $\mathbb{P}(X_k = 0)$ can be written as

$$\mathbb{P}(X_k = 0) = \mathbb{P}(X_k = 0 | \tau \leq k) \mathbb{P}(\tau \leq k) + \mathbb{P}(X_k = 0 | \tau > k) \mathbb{P}(\tau > k).$$

Now if we can prove that for $k \geq \zeta n$, $\mathbb{P}(\tau > k)$ goes to zeros as n tends to infinity, then we would have

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(X_k = 0) &= \lim_{n \rightarrow \infty} \mathbb{P}(X_k = 0 | \tau \leq k) \mathbb{P}(\tau \leq k) + \mathbb{P}(X_k = 0 | \tau > k) \mathbb{P}(\tau > k) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(Y_k = 0 | \tau \leq k) \mathbb{P}(\tau \leq k) + \lim_{n \rightarrow \infty} \mathbb{P}(X_k = 0 | \tau > k) \mathbb{P}(\tau > k) \\ &= \lim_{n \rightarrow \infty} \left(\frac{1}{2}\right)^n \mathbb{P}(\tau \leq k), \end{aligned}$$

which proves the equation (39). Note that $\mathbb{P}(\tau \leq k)$ tends to 1 as k approaches infinity. Therefore to complete the proof it remains to show that

$$\lim_{n \rightarrow \infty} \mathbb{P}(\tau > \zeta n) = 0. \quad (40)$$

For $1 < j \leq n$, suppose τ_j represents the first time that the j^{th} entries of X_t and Y_t become the same.

Thus we have

$$\mathbb{P}(\tau > \zeta n) \leq \sum_{j=1}^n \mathbb{P}(\tau_j > \zeta n) = n \mathbb{P}(\tau_1 > \zeta n).$$

Suppose that after i steps the first entries of X_i and Y_i are not the same and let r_i and the set U be as defined previously. Let $\rho < \frac{1}{2}$. The first entry of s_{i+1}^x is equal to one with probability ρ . Now due to the fact that $\rho < \frac{1}{2}$, less than $\frac{n-r_i}{2}$ entries of s_{i+1}^x which are not from U are equal to one with a probability more than $\frac{1}{2}$. This means that the first entry of s_{i+1}^y is equal to zero with a probability more than $\frac{1}{4}$. Thus, the first entries of s_{i+1}^x and s_{i+1}^y are not the same with a probability more than $\frac{\rho}{4}$. A similar approach for the case $\rho \geq \frac{1}{2}$ shows that there is a positive probability p independent from n and i such that the first entries of s_{i+1}^x and s_{i+1}^y differ, i.e., the first entries of X_{i+1} and Y_{i+1} are the same. Thus we have

$$\lim_{n \rightarrow \infty} \mathbb{P}(\tau > \zeta n) \leq \lim_{n \rightarrow \infty} n \mathbb{P}(\tau_1 > \zeta n) \leq \lim_{n \rightarrow \infty} n(1-p)^{n\zeta} = 0.$$

This completes the proof. ■

ACKNOWLEDGMENT

The authors would like to thank Professor G. D. Forney for his valuable comments and suggestions and Mr. R. Farhodi for his comments about proof of theorems.

REFERENCES

- [1] C. E. Shannon, A mathematical theory of communications, *Bell Systems Technical Journal*, vol. 27, pp. 379-429, 1948.
- [2] R. M. Fano, *Transmission of Information*, The M.I.T. Press, Cambridge, 1961.
- [3] R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Transactions Information Theory*, vol. 11, no. 1, pp. 3-18, Jan. 1965.
- [4] Y. Polyanskiy, H. Vincent Poor, S. Verdú, "Channel Coding Rate in the Finite Blocklength Regime," *IEEE Transactions Information Theory*, vol. 56, issue 5, pp. 2307-2359, May 2010.
- [5] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley and Sons Inc. New York, NY, USA 1968, p. 204.
- [6] M. Mezard and A. Montanari, *Information, physics, and computation*, Oxford University Press, USA, 2009, pp. 105-128.
- [7] A. Barg and G. D. Forney, "Random codes: Minimum distances and error exponents," *IEEE Transactions Information Theory*, vol. 48, no. 9, pp. 2568-2573, Sept. 2002.
- [8] G. Poltyrev, "Bounds on the decoding error probability of linear binary codes via their spectra," *IEEE Transactions Information Theory*, vol. 40, no. 4, pp. 1284-1292, Jul. 1994.
- [9] R. G. Gallager, "Low density parity check codes," *IRE Transactions Information Theory*, vol. IT-8, pp. 21, Jan. 1964.
- [10] D. J. C. MacKay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *IEE Electronics Letters*, vol. 33, no. 6, pp. 457-458, Jul. 1997.
- [11] F. J. Vazquez-Araujo, M. Gonzalez-Lopez, L. Castedo, and J. Garcia-Frias, "Capacity approaching low-rate LDGM codes," *IEEE Transactions Communications*, vol. 59, no. 2, pp. 352-356, Feb 2011.
- [12] J. Garcia-Frias and Z. Wei, "Approaching Shannon performance by iterative decoding of linear codes with low-density generator matrix," *IEEE Communications Letters*, vol. 7, no. 6, pp. 266-268, June 2003.
- [13] H. Chun-Hao and A. Anastasopoulos, "Capacity-achieving codes with bounded graphical complexity and maximum likelihood decoding," *IEEE Transactions Information Theory*, vol. 56, no. 3, pp. 992-1006, March 2010.
- [14] M. Luby, "LT codes," in *Proc. IEEE Symposium on Foundations of Computer Science*, pp. 271-280, 2002.
- [15] A. Shokrollahi, "Raptor codes," *IEEE Transactions Information Theory*, vol. 52, no. 6, pp. 2551-2567, June 2006.
- [16] A. Dembo and O. Zeitouni, *Large Deviation Techniques and Application*, Springer, 2009.
- [17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, 1991, p. 25.
- [18] C. Cooper, "On the rank of random matrices," *Random Structures Algorithms*, vol. 16, no. 2, pp. 209-232, Feb. 2000.

- [19] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 13-30, 1963